INTRODUCTION TO VOLUME 7: LANGUAGE TESTING AND ASSESSMENT   Au1

This volume addresses the broad theme and specific topics associated with current thinking in the field of language testing and assessment. Interdisciplinary in its nature, language testing and assessment build on theories and definitions provided by linguistics, applied linguistics, language acquisition and language teaching as well as on the discipline of testing, measurement and evaluation. Language testing uses these disciplines as foundations for researching, theorizing and constructing valid language tools for assessing and judging the quality of language used.

The field of language testing is therefore viewed as consisting of two major components; one focusing on the 'what', referring to the constructs that need to be assessed, also known as 'the trait', and the component pertaining to the 'how', known also as 'the method', which addresses the specific procedures and strategies used for assessing the 'what'. Traditionally, the trait has been defined by the language fields as these provide the essential elements that language testers needed to incorporate in the process of creating tests. The 'how', on the other hand, is derived mostly from the field of testing and measurement   Au2 which has, over the years, developed a broad body of theories, research, techniques and practices about testing and assessment. Language testers incorporate these two areas to create the discipline of language testing and assessment, a broad field which includes theories, research and applications; it has its own research, publications, conferences and the journals *Language Testing* and *Language Assessment Quarterly* where many of these publications appear.

An examination of the developments in the language testing and assessment field since the 1960's reveals that its theories and practices have always been closely related to definitions of language proficiency. Matching the 'how' of testing with the 'what' of language, uncovers several periods in the development of the field, with each instantiating different notions of language knowledge and the specific measurement procedures that go with them. Thus, discrete-point testing viewed language as consisting of lexical and structural items so that the language test of that era presented isolated items in objective testing procedures. In the integrative era, language tests tapped integrated and discoursal language; in the communicative era tests aimed to replicate interactions

among language users utilizing authentic oral and written texts; and in the performance testing era language users were expected to perform tasks in well-defined and real life contexts. Alternative assessment was a way of responding to the realization that language knowledge is a complex phenomenon that no single procedure can be expected to capture and it therefore requires multiple and varied assessment procedures that complement one another.

While the 'what' to 'how' trajectory for the development of tests is intuitively plausible, extensive work in the past decade points to a less overt but highly influential dynamic in another direction, that is, the pivotal roles that tests play in societies in shaping the definitions of language, in affecting learning and teaching and in maintaining and creating social classes. That means that contemporary assessment research perceives its obligations also to examine the close relationship between methods and traits in broader contexts which focus on how language tests interact with these societal factors, given their enormous power. In other words, as language testers seek to develop and design methods and procedures for assessment (the 'how') they also become mindful not only of the emerging insights regarding the trait (the 'what'), its complexity, its multiple facets and dimensions, but also of the societal role that language tests play, the power they hold and their central functions in education, politics and society.

In terms of the interaction of society and language, it is evident that changes are currently occurring in the broader contexts and spaces in which language testing takes place. It is realized that language testing is not occurring in homogenous, uniform and isolated contexts but rather in diverse, multilingual and multicultural societies and thus posing new challenges and questions with regards to what it means to know language(s) in education and society. For example, different meanings of language knowledge may be associated with learning foreign languages, second languages, language by immersion, heritage languages, languages of immigrants arriving to new places with no knowledge of the new languages and languages of those defined as 'trans-nationals'. Knowing the English language, the current world's lingua franca is different than knowing other languages and language in classrooms and schools may be different than language at the workplaces or at communities where bi- or multilingual patterns are the norm. Each of these contexts may require different and varied theories of language knowledge and hence different definitions and applications of measuring these proficiencies.

In other words, the languages currently being used in different societies in different contexts, no longer represent uniform constructs as these vary from one place to another, from one context to another, creating different language patterns, expectations and goals and often

resulting in hybrids and fusions, especially with regards to English. Such dynamic linguistic phenomena pose challenging problems to language testers. What is/are the language(s) that needs to be assessed? Where can it be observed in the best ways? Is it different at home, schools, classrooms and the workplace? Should hybrids and fusions be assessed and how? Can levels of languages even be defined? How should language proficiency be reported and to whom? What is 'good language'? Does such a term even apply? Who should decide how tests should be used? Do testers have an obligation to express their views about language and testing policy? What is the responsibility of testers to language learning, language use in classrooms and communities? How can ethical and professional behaviours with regards to tests be maintained? These are some of the questions that language testers are currently pre-occupied with. Language testers are not technicians that just invent better and more sophisticated testing tools, but rather they are constantly in search for and concerned about the 'what' and its complex meanings. The unique aspect of 'language' testing is that, it is an integral part of a defined discipline, that of 'language'. In that respect, language testers and the field of language testing are different than the field of 'general testing'. Language testers are confined to a specific discipline and are therefore in constant need of examining such questions and the validity of the tools against these disciplines.

The concern of language testers in the past decade about the use of tests, their political, social, educational and ethical dimensions, made the field even more complex and uncertain and in need of new questions and debates. The current era can be described as the era of uncertainty where questions are being raised about the meaning of language and the ways of measuring this complex and dynamic language trait. At the same time, it is an era of ever more compelling need to ensure that these tests are reliable and valid, where validity includes protection and guarding of the personal rights of others and also to ensure positive washback on learning by addressing the diverse communities in which the tests are being used. Thus, the current era is not only concerned with a broader and more complex view of what it means to know a language, not only with innovative methods of testing and assessing these complex constructs, but also with how these tests can be more inclusive, more democratic, more just, more open, more fair, more equal and less biased. Even within the use of traditional large scale testing, the field is asking questions about tests' use—why test, who benefits, who loses, what are the impacts and consequences on definition of language, on people, education, society and language policies? Tests are not viewed as innocent tools but rather as instruments that play central roles for people, education and societies. Language testers therefore are asked to deal with broader issues: to examine the uses of tests in

the complex multilingual and multicultural societies where these tests are used not only as naïve measurement tools but also as powerful educational, societal and political tools.

This is the conceptual premise of this volume on Language and Assessment of the *Encyclopedia of Language and Education*. It aims to cover and uncover the multiple versions and perspectives of the 'what' of languages, the multiple ways of assessing the 'what', the approaches developed for assessment, given the multiplicity of languages used by diverse groups of learners in different contexts as well as the societal roles and obligations of language testing to be accountable and to ensure ethicality and professionalism. A special focus is given in this volume to the multilingual and diverse contexts in which language testing and assessment are currently anchored, and the difficult tasks of 'doing testing' in this complex day and age.

Accordingly, the first part of the volume addresses the 'what' of language testing and assessment. It no longer divides language into neat and clear-cut skills of reading, writing, speaking and listening, but rather examines the 'what' of language in the diverse contexts in which it is used. Rather than proposing one uniform way of defining the language construct, the chapters in Part 1 present language from multiple perspectives. It begins with a chapter by Alister Cumming who reviews research and practices of language assessment from the perspectives of oral and literate modes of communication and their meanings in relation to language competencies, language learning and multimodalities, noting that language assessment needs to be informed and extended by multipe forms of evidence in relation to educational purposes as well as diverse societies. Rama Mathew surveys developments in language assessment from the perspective of multilingual competence as manifested in India. She shows how the legitimacy of multilingual reality in many places in the world today and the tasks of maintaining this reality call for different thinking about language assessments. She demonstrates this through a thorough survey of multilingual and multidialect tests in assessing English in India and the various assessment issues that are relevant to these realities. Heidi Byrnes focusses on content as part of language proficiency as language and content are closely embedded in one another especially on the advanced levels of proficiency. She surveys assessment practices beyond traditional views of formal features separate from knowing a content areas. By using a Hallidayan approach to texts and knowledge, she shows how assessment can be interpreted as part of sophisticated meanings of texts and knowledge related to the kind of textual environment necessary for handling content and granting priority to various elements of texts and their conribution to content. These topics are discussed within multilingualism, integrating L2 into existing mainstream curricula,

diverse content expectations and second language literacy in diverse professional contexts as part of global literacies.

The perspective of models of Communicative Language Ability in defining language is taken by Jim Purpura who uses it as the basis for test development. He surveys the different theoretical models and the tests which were developed based on these models. And argues that they represent targets of assessment that can be adapted to a range of test purposes and contexts especially in the areas of grammar and its meanings, consisting of both grammatical and pragmatic knowledge. Accordingly, tests which are developed based on these theories can help in better understanding the components underlying communicative language ability as well as to provide useful diagnostic information to learners. Kieran O'Loughlin examines language from the angle of the workplace as part of language for specific purposes where distinctions are made between academic purposes and occupational purposes relevant to professional duties. He provides an historical review and current practices of the different performance-based tests related to the 'real world' of functions and tasks for employment in a number of professional areas as part of global demands. He is sceptical on the future of these tests given the spread of large scale standardized tests. Stansfield and Winke provide a somewhat different perspective of the construct by re-visiting language aptitude. They re-define the construct of language aptitude by expanding its meaning to include also aspects of second language learning such as diagnosis and treatment of L2 learning problems, informing curricular design and examining how factors of working memory are related to L2 learning as part of cognitive abilities. They survey the types of aptitude tests that are in line with these new theoretical and complex constructs and levels of predictability, but raise questions about the validity of these tests and their uses. Together, these six chapters provide an expanded and multiple perspective view of the language construct and the type of assessments practices that are associated with these perspectives. There is clearly no one way to define language as these chapters demonstrate since the definitions of language cannot be detached from the diverse contexts and purposes in which it is used.

The second part of the volume addresses the diverse methodological issues that language testers face in assessing the complex construct of language, that is, the 'how'. The chapters demonstrate the sophisticated issues and deliberations as well as specific procedures used for assessing language. The first chapter is by Janna Fox who reviews developments and procedures of alternative assessment. Fox expands the theoretical issues of alternative assessment not only by providing a longer list of 'alternatives' but especially in challenging the issue of whether alternative assessment represents a real paradigm shift or just

added procedures that preserve traditional ways of testing. She further exapands the notion by including different ways of thinking of testing in the alternatives modes in the direction of test accommodations, dynamic assessment and shows how such approaches support a more ethical, democratic, equitable testing in line with values, individual diversity and assessment culture. One of the dominant cases of alternative assessment is that of task and performance, issues that Jill Wiggleworth reviews in her chapter which focusses on the very tasks that are designed to measure learners' productive language skills through performances which are related to real world contexts (e.g. the workplace). She surveys the vast research literature demonstrating the value of certain performance tests, the effect of certain variables on quality of tasks such as difficulty levels, cognitive demands, type of discourse they produce as well as the extent to which they indeed represent 'real life'. In continuation with the variety of assessment methods Carol Chapelle delineates the new and current methods of utilizing technology in language assessment—Computer-Assisted Testing (CAT) by reviewing tests using Micro computers and the Internet, not only in terms of their greater efficiency but also in terms of the serious problems they pose. She surveys research on multimedia testing and their effects on learners, specific skills such as listening, natural language processing technologies and the usability of assessing written and spoken language. Issues of access to infrastructure, costs, the intersections with construct validity and training are brought up along with questions as to whether computerized testing has been evolutionary rather than revolutionary.

While the debates on the appropriate methods of assessment are taking place, large scale testing continues to be administered with even more force than ever before by governments and educational systems world wide. In schools, tests are used to monitor students' progress through standardized tests and for diagnostic purposes; at college and university levels they are used for screening and selection of applicants and measuring achievements for accountability purposes and ensuring equal opportunities. Anthony Kunnan reports and discusses these issues and raises questions about the advantages of uniformity of these tests in the sake of fairness. He reviews the history of large scale testing and provides safeguards as to fairness in the form of descriptive test information, codes of practices, practical matters of test design and psychometric qualities.

Methods such as the Common European Framework and other standards have been receiving major attention and dominance in the past decade, especially with regards to their effects on the definitions of language and language policy. Glenn Fulcher describes the multiple methods of examining the quality of language via rating scales, standards, benchmarks, band levels and guidelines and their often overlapping

meanings. These methods are becoming the main paradigms with strong effects on language learning and teaching, beyond being just criteria for evaluation of language quality. By presenting these descriptions he shows their advantages but especially their flaws in terms of validity of progression, equivalence across languages, hierarchies, false claims and effects on definitions of language. He points to the inherent difficulty in trying to relate test scores to direct ability to perform on non-test tasks.

The field of psychometrics has gone through major changes in attempting to accommodate the more complex tests and tasks so they will pass criteria of reliability, validity and ethicality. The chapter by Xiaoming Xi provides a comprehensive examination of these issues and updated methods of test validation. She shows how advances in validity benefit from progress in educational measurement, psychometrics and statistics, qualitative methods, discourse analysis, cognitive psychology as well as introspective methods about tasks complexity. Anne Lazaraton introduces new ways of utilizing qualitative methods for designing, describing and validating language tests, a topic that is obtaining acceptance and legitimacy within the field of language testing, especially given the limitations of traditional statistical methods. She demontrates how qualitative methods can provide indication of the quality of tests both on the process and the product levels. The chapter co-authored by Micheline Chaloub-Deville and Craig Deville examines the common psychometric methods that are used in the field through a careful review and analysis of language testing research as reported in highly regarded testing journals. They show the multiple and varied methods which are used in testing research.

Meg Malone introduces the topic of training and teaching about language testing given the vast amount of knowledge available today so that testers can make informed decisions throughout the assessment process about test development, scoring, interpretation, selection and administration of tests. She introduces the term 'assessment literacy' to refer to the required knowledge about testing and its multiple interpretations. Another new topic relates to the emerging field of corpus linguistics. In the chapter by Linda Taylor and Fiona Baker they illustrate how the field of corpus linguistics has become an important source of relevant data for the language testing field in obtaining a more accurate and content valid picture of the very language that needs to be assessed and its uses on tests in different contexts based on scientific and empirical methods. Together the chapters in Part 2 provide multiple considerations as well as specific methods that language testers are applying in developing language tests including problems and future issues.

Part 3 of the volume addresses issues of language testing embedded within educational systems and contexts, where language tests are

widely used and represent a common practice. It is in the educational system that tests and various assessment methods serve as major tools for assessing language for learning and teaching, making decisions about programmes, teachers and learners as well as creating changes leading to school reforms and intended and un-intended washbacks in classrooms and schools on local, national and international levels. Pauline Rae-Dickins opens Part 3 with a chapter which focuses on a rather overlooked area in relation to external high stakes test. She makes the distinctions between assessment *of* learning, that is focused on achievement and summative in orientation, and assessment *as* learning, that is formative in purpose providing feedback to learners so that they can improve their learning. Yet, she points to the ample progress in the latter area in the past decade and surveys these studies, especially with regards to wide uses that teachers make with different assessment tools in the classroom for feedback and learning leading to effective instructional methods. Another new topic that received recognition in the past decade both in general educational testing and in language testing is that of Dynamic Assessment. Lantolf and Poehner introduce this topic in the context of *language* testing and show how by applying Vygotsky's sociocultural theories, testing and learning can become closely connected, leading to more effective learning whereby tests are not just separate assessment tools but are rather embedded *in* learning and can lead to improved learning. Another development of the past decade is the increased attention to culture of assessment as part of ongoing effective learning and teaching in schools. Ofra Inbar reviews studies that address this issue of 'testing culture' and shows how assessment has a new role as part of effective and beneficial learning, showing how the continuous use of thebroad information provided by good assessment can produce can meaningful changes in the school organization; testing then becomes an integral part of school culture.

It is the recognition that current schools are diverse in terms of students' languages and cultural backgrounds that requires different approaches to assessment especially where immigrants and indigenous students are faced with situations whereby they are not proficient in the language of the instruction in schools and are engaged in language acquisition while attempting to acquire school content. Doing tests in the dominant school languages is a difficult goal that has major implications for academic success. Several chapters in Part 3 address these issues. Constant Leung and Jo Lewkowicz provide a most comprehensive overview of the types of assessment procedures used in diverse multilingual and plurilingual communities in the context of both second language assessment designed to measure language development of linguistic minority students where the language is the predominant the society majority language as well as in contexts where the

second/foreign language assessment are designed to measure language development of learners of a language in diverse contexts. They note that new developments are indicative of more progressive views recognizing the value of a people's proficiencies in different languages. Cath Rau discusses assessment strategies for indigenous populations in schools in places where indigenous groups make up the big part of the population, as in the case of the Maöri and other groups in New Zealand. She surveys descriptions of a number of strategies used to practice testing in more fair ways incorporating existing language knowledge. Test accommodations refer to strategies used for language learners to assess their content knowledge while compensating for lack of language knowledge in order to create more fair testing conditions for those for whom the language of assessment is their second language in relation to native speakers in relation to academic subjects. Jamal Abedi reviews the extensive research that has been conducted in the past decade on the topic examining the effective accommodations for language learners, mostly in the context English language learners in the USA. He brings evidence from research about different types of accommodations in content areas such as Mathematics while being critical on the uses of some accommodations that have no empirical bases.

Issues of washback and impact of large scale testing on learning have stimulated ample research and writings in the past decade. The chapter by Liying Cheng surveys a large number of studies that documented the kind of effects and impacts tests have on learning, teaching and the curriculum, based on empirical studies conducted in many educational systems world-wide. It is evident from this chapter how this topic has advanced in the past decade with empirical studies so that test washback is considered today a part of construct validity and incorporated in developments of large scale tests. Geoff Brindley demonstrates how language tests are used by governments to reform educational systems and points to the serious problems and issues related to such practices of relying exclusively on tests for educational reforms. Alison Bailey addresses the methods and techniques used for assessing the language of young learners in schools pointing at the different strategies of these tests compared to those for adults. This topic is gaining major attention nowadays with the growing number of young learners learning English around the world. Taken together, the chapters in Part 3 of the volume cover a wide range of topics related to broad issues of language assessment in education as they face challenges given the changing realities of diverse populations and the realization that language tests play a major role in learning as well as in school reforms.

The fourth and concluding part of this volume addresses societal, political, professional and ethical dimensions of tests; a topic that has been a major concern in the language testing field in the past decade.

Each of the four chapters addresses different aspects of these dimen-
sions. The chapter by Kate Menken illustrates how national language
tests, especially those administered by government initiatives (e.g. the
No Child Left Behind mandate in the USA) affect language policy in
schools and societies and deliver specific messages about the signifi-
cance and insignificance of certain languages and language instruction
policies. She shows how language testing and language policy are
closely connected as language tests have a greater effect than is viewed
on the surface. This is especially relevant in societies which include
learners for whom the language of the test affects their ability to per-
form academically. She reviews cases where language tests become
in fact the de facto language policies leading to multiple ethical pro-
blems. Tim McNamara explains the need for taking into account the
socio and political dimensions of tests versus the structualist and psy-
chometric dimensions which have dominated the field. In his chapter,
he surveys various social theories of linguistics as important input to the
field of language testing with special attention to the work of Messick
who described the values and consequences of tests as part of construct
validity. He surveys the theoretical and practical developments of stu-
dies and cases where language tests are used unjustly such as in the
cases of citizenship and employment and asylum seekers. Alan Davies,
who has written extensively on the ethical dimensions of tests and
especially on the professional aspects related to ethicality, addresses
these issues by covering the developments in the field and showing how
the Code of Ethics and Code of Practice, developed by the language
testing profession via the International Language Testing Association,
can lead to more ethical tests use. Yet, he warns against the use of such
codes as face saving devices while overlooking the real commitment to
ethics that is instrumental for the profession itself, for its stakeholders
and the rights of individual test takers.The final chapter is by Bernard
Spolsky who examines the history, present and future of the field of
language testing, providing guidance and material for future vision.
He surveys the history of the field with its advances as well as the
ample questions and? uncertainties that emerge and that need to be
addressed in the future, while pointing to the contradictions, problems
and difficulties of measuring and assessing such a complex construct as
'language'. He ends the chapter by stating that he remains sceptical
given the role of industrial test makers, computerizing tests, reducing
multidimensional profiles into uniform scales, and the educational sys-
tems that continue to interpret test scores as if they are meaningful. At
the same time, he notes that he expects the quality research that has
been conducted in the field of language testing to continue especially
in the areas of the nature of language proficiency and the demonstration
of possible ways to assess its relevance to defined social purposes.

The editor would like to thank each and every author of these chapters which together make up a most valuable contribution to the field of language testing and to applied linguistics. The authors selected to write these chapters are among the most distinguished scholars and leaders in the field of language testing. The chapters herein reveal that the language testing field is dynamic, striving and vital. It is clear from these chapters that the field of language testing raises important and deep questions and does not overlook problems, difficulties, contradictions, malpractices and new societal realities. While viewed by some as a technical field, this volume convincingly demonstrates that language testing and assessment is above all a scholarly and intellectual field that touches the essence of languages and their meanings. The need to get engaged in testing and assessment forces testers to face these issues head-on and attempt to deliberate creative and thoughtful solutions.

Finally, special personal and deep thanks to Caroline Clapham who in her 1997 volume on Language Testing in the first edition of the Encyclopedia of Language and Education set the foundations and grounds for the field in such a comprehensive way that it was possible to expand and create this very comprehensive and stimulating volume.

Elana Shohamy

Comp. by: MGrahalatshmi   Date:5/4/07   Time:18:15:32   Stage:First Proof
File Path://spiina1001z/womat/production/PRODENV/0000000005/0000001817/
0000000016/0000590828.3D   Proof by: QC by:

Author Query Form

| Encyclopedia of Language and Education |
| Chapter - 11963 |

_____

Query Refs.	Details Required	Author's response
AU1	Please provide the affiliation and corresponding details of the author.	
AU2	Elana, I deleted this because it seems to suggest that the field of testing is dynamic, but the language fields are not?	

Springer